Recap on convex optimization

Carlos Cotrini
Department of Computer Science
ETH Zürich

October 12, 2025

1 Introduction

Definition 1. *Let $f, g_1, g_2, \dots, g_m, h_1, h_2, \dots, h_n$ be functions with domain in $\mathbb{R}^d$ and codomain in $\mathbb{R}$. A* constrained optimization problem *is a problem of the form*

$$\begin{aligned} \min_{w \in \mathbb{R}^d} \quad & f(w) && (1)\\ s.t. \quad & g_i(w) = 0, \quad \textit{for } i \leq m, && (2)\\ & h_j(w) \leq 0, \quad \textit{for } j \leq n. && (3) \end{aligned}$$

The equations in (2) and (3) are called the constraints*. A point $w \in \mathbb{R}^d$ is* a feasible solution *if it satisfies all the constraints and is* an optimal solution *if $f(w) \leq f(w')$ for any other feasible point $w' \in \mathbb{R}^d$.*

Definition 2. *A* convex optimization problem *is a constrained optimization problem where $f, g_1, \dots, g_m, h_1, \dots, h_n$ are convex and the set of feasible solutions is convex.*

2 Illustrative example

Consider a drone located at a fixed position (x_0, y_0, z_0) in the Euclidean space and a person standing on a flat disk of radius r whose center is at the origin. The person wants to get as close as possible to the drone, but cannot leave the disk, otherwise it falls in the direction $(0, 0, -1)$. We can formulate this as a convex constrained optimization problem.

$$\begin{aligned} \min_{(x,y,z) \in \mathbb{R}^3} \quad & \|(x, y, z) - (x_0, y_0, z_0)\|^2 && (4)\\ s.t. \quad & z = 0, && (5)\\ & x^2 + y^2 \leq r. && (6) \end{aligned}$$

Let

$$\begin{aligned} f(x, y, z) &:= \|(x, y, z) - (x_0, y_0, z_0)\|^2, && (7)\\ g(x, y, z) &:= z, \text{ and} && (8)\\ h(x, y, z) &:= x^2 + y^2 - r. && (9) \end{aligned}$$

Observe that

$$\nabla f(x,y,z) = (x - x_0, y - y_0, z - z_0), \tag{10}$$
$$\nabla g(x,y,z) = (0,0,1), \text{ and} \tag{11}$$
$$\nabla h(x,y,z) = (2x, 2y, 0). \tag{12}$$

Exercise 1. *Convince yourself that $\nabla g(x,y,z)$ is always perpendicular to the disk surface and that $\nabla h(x,y,z)$ is perpendicular to the disk's edge, but only when (x,y,z) is on the disk's edge. You do not need to formally prove this.*

Exercise 2. *For the following cases, give an equation in terms of the gradients above that defines a necessary condition for an optimal solution for Problem 4. You do not need to formally prove your answers.*

1. *The drone is lying on the disk.*
2. *The drone is not on the disk, but inside the cylinder $x^2 + y^2 \leq r$.*
3. *The drone is outside the cylinder $x^2 + y^2 \leq r$.*

The solutions are

1. $\nabla f(x,y,z) = 0$.
2. $\nabla f(x,y,z) + \lambda\, g(x,y,z) = 0$, for some $\lambda \in \mathbb{R}$.
3. $\nabla f(x,y,z) + \lambda\, g(x,y,z) + \alpha\, h(x,y,z) = 0$, for some $\lambda, \alpha \in \mathbb{R}$, with $\alpha > 0$.

Notice that we can rewrite them into a more uniform but slightly more complicated way as follows:

1. $\nabla f(x,y,z) + \lambda\, g(x,y,z) + \alpha\, h(x,y,z) = 0$, for some $\lambda, \alpha \in \mathbb{R}$, with $\lambda = 0, \alpha = 0$.
2. $\nabla f(x,y,z) + \lambda\, g(x,y,z) + \alpha\, h(x,y,z) = 0$, for some $\lambda, \alpha \in \mathbb{R}$, with $\lambda \neq 0, \alpha = 0$.
3. $\nabla f(x,y,z) + \lambda\, g(x,y,z) + \alpha\, h(x,y,z) = 0$, for some $\lambda, \alpha \in \mathbb{R}$, with $\lambda \neq 0, \alpha > 0$.

We see then that we can unify all these conditions into one single necessary condition:

$$\nabla f(x,y,z) + \lambda\, \nabla g(x,y,z) + \alpha\, \nabla h(x,y,z) = 0, \text{ for some } \lambda, \alpha \in \mathbb{R}, \text{ with } \alpha \geq 0. \tag{13}$$

By linearity of the gradient operator, we can rewrite this as

$$\nabla_{x,y,z}\left(f(x,y,z) + \lambda\, g(x,y,z) + \alpha\, h(x,y,z)\right) = 0, \text{ for some } \lambda, \alpha \in \mathbb{R}, \text{ with } \alpha \geq 0. \tag{14}$$

Observe that the gradient is only with respect to the variables x, y, and z and not with respect to λ or α. Surprisingly, one can demonstrate that Equation 14 holds for *any* convex constrained optimization problem.

Exercise 3. *Assume that the drone is at a fixed position (x_0, y_0, z_0) and that the person's coordinates are restricted so that $y = 0$, $z = 0$, and $0 \leq x \leq 1$. The person wants to be as close as possible to the drone. Formulate this as a convex optimization problem of the form (1). Then convince yourself that an optimal solution must fulfil the equation below. You do not need to formally justify this.*

$$\nabla_w \left(f(w) + \sum_{i \leq m} \lambda_i g_i(w) + \sum_{j \leq n} \alpha_j h_j(w) \right) = 0, \text{ with } \alpha_j \geq 0. \tag{15}$$

3 Lagrangians

For many simple constrained problems, Equation 15 is already good enough to narrow down the number of possible solutions to a number that can be manually inspected to find an optimal solution. Because of its usefulness, the argument of the gradient in the left-hand side of Equation 15 has become an important concept in optimization.

Definition 3. *The* Lagrangian *of a constrained optimization problem of the form (1) is*

$$\mathcal{L}(\lambda_1, \ldots, \lambda_m, \alpha_1, \ldots, \alpha_n, w) = f(w) + \sum_{i \leq m} \lambda_i g_i(w) + \sum_{j \leq n} \alpha_j h_j(w). \tag{16}$$

For convenience, we often abbreviate $\lambda_1, \ldots, \lambda_m$ with λ and $\alpha_1, \ldots, \alpha_n$ with α. We also often write $\alpha \geq 0$ to denote $\alpha_j \geq 0$, for any $j \leq n$.

Lemma 1. *Any optimal solution w^* of a convex optimization problem of the form (1) must fulfil*

$$\begin{aligned} \nabla_w \mathcal{L}(\lambda, \alpha, w) &= 0, & (17)\\ g_i(w) &= 0, \text{ for } i \leq m, & (18)\\ h_j(w) &\leq 0, \text{ for } j \leq n, \text{and} & (19)\\ \alpha_j &\geq 0, \text{ for } j \leq n. & (20) \end{aligned}$$

The proof is usually done in convex optimization courses.

4 Duals

In many interesting convex optimization problems, solving the equations from Lemma 1 is intractable. One alternative to this issue is to compute the *dual* optimization problem associated to the original problem. Solving the dual can be easier, and if a condition called *strong duality* is met, then one can calculate an optimal solution for the original problem from a dual's optimal solution.

We now derive the dual problem from a generic convex optimization problem.

Exercise 4. *Show that if w is feasible and $\alpha \geq 0$, then:*

$$\begin{aligned} \mathcal{L}(\lambda, \alpha, w) &= f(w) + \sum_{i \leq m} \lambda_i g_i(w) + \sum_{j \leq n} \alpha_j h_j(w) & (21)\\ &= f(w) + \sum_{j \leq n} \alpha_j h_j(w) & (22)\\ &\leq f(w). & (23) \end{aligned}$$

Therefore, for any λ, $\alpha \geq 0$, and w feasible,

$$\inf_w \mathcal{L}(\lambda, \alpha, w) \leq \mathcal{L}(\lambda, \alpha, w) \leq f(w) \leq f(w^*). \tag{24}$$

Observe that the infimum is over all possible values for w, not necessarily those that are feasible. Finally, since this holds for any λ and any $\alpha \geq 0$, we can conclude that

$$\max_{\lambda,\, \alpha \geq 0} \inf_{w} \mathcal{L}\left(\lambda, \alpha, w\right) \leq f(w^*). \tag{25}$$

Let $\theta(\lambda, \alpha) := \inf_w \mathcal{L}\left(\lambda, \alpha, w\right)$. Observe that θ is the infimum of a linear combination of convex functions. One can demonstrate that such a function is also convex, so θ is a convex function. Equation 25 shows that

$$\theta(\lambda^*, \alpha^*) \leq f(w^*), \tag{26}$$

where (λ^*, α^*) is an optimal solution of the following optimization problem.

$$\max_{\lambda \in \mathbb{R}^m,\, \alpha \in \mathbb{R}^n} \quad \theta(\lambda, \alpha) \tag{27}$$

$$s.t. \quad \alpha \geq 0. \tag{28}$$

Definition 4. *For a convex optimization problem of the form (1), its* dual *is the convex optimization problem of the form (27).*

Observe how the dual trades a simpler set of constraints for a more complex objective function. This trade-off is very helpful when computing certain support-vector machines, as θ can be simplified into an expression easier to calculate, yielding a simpler convex optimization problem.

5 Strong duality

Definition 5. *A convex optimization problem satisfies* strong duality *if $\theta(\lambda^*, \alpha^*) = f(w^*)$, where w^* is an optimal solution of the problem and (λ^*, α^*) is an optimal solution of the dual.*

Strong duality is a desirable property. We show how it yields a way to compute an optimal solution for a convex optimization problem from an optimal solution for its dual. However, deciding strong duality without solving the convex optimization problem or its dual is hard. Fortunately, several sufficient conditions have been established for strong duality. A simple one is Slater's condition, which requires the existence of a feasible solution that fulfils all inequality constraints strictly.

Definition 6. *A convex optimization problem of the form (1) satisfies* Slater's condition *if there is a feasible solution w^o such that $h_j(w^o) < 0$, for $j \leq n$.*

Lemma 2. *If a convex optimization problem satisfies Slater's condition, then it satisfies strong duality.*

The converse is not true.

Lemma 3. *If a convex optimization problem satisfies strong duality then the following hold for any optimal solution w^*.*

$$f(w^*) = \mathcal{L}\left(\lambda^*, \alpha^*, w^*\right), \ \textit{where } (\lambda^*, \alpha^*) \textit{ is an optimal solution for the dual.} \tag{29}$$

$$\alpha_j h_j(w^*) = 0, \ \textit{for } j \leq n. \tag{30}$$

The condition in Equation 30 is called *complementary slackness.*

Proof. Let w^* be an optimal solution, then

$$\begin{aligned} f(w^*) &= \theta(\lambda^*, \alpha^*) && (31)\\ &= \inf_w \mathcal{L}(\lambda^*, \alpha^*, w) && (32)\\ &\leq \mathcal{L}(\lambda^*, \alpha^*, w^*) && (33)\\ &\leq f(w^*). && (34) \end{aligned}$$

This implies that

$$f(w^*) = \mathcal{L}(\lambda^*, \alpha^*, w^*), \tag{35}$$

which is one of the things we wanted. Now, expanding the definition of the Lagrangian yields that

$$f(w^*) = \mathcal{L}(\lambda^*, \alpha^*, w^*) = f(w^*) + \sum_{i \leq m} \lambda_i g_i(w^*) + \sum_{j \leq n} \alpha_j h_j(w^*). \tag{36}$$

Hence,

$$0 = \sum_{i \leq m} \lambda_i g_i(w^*) + \sum_{j \leq n} \alpha_j h_j(w^*). \tag{37}$$

However, $g_i(w^*) = 0$, for $i \leq m$, and $\alpha_j h_j(w^*) \leq 0$, for $j \leq n$, so it must be the case that

$$0 = \alpha_j h_j(w^*), \text{ for } j \leq n. \tag{38}$$

This concludes the proof. □

6 How to solve convex optimization problems

We now present some strategies for solving convex optimization problems using the lemmas above.

One simple strategy is to verify if Slater's condition is fulfilled. If that is the case, then one proceeds to solve Equations 15, 2, 3, 28, and 30:

$$\begin{aligned} \nabla_w \mathcal{L}(\lambda, \alpha, w) &= 0 && (39)\\ g_i(w) &= 0, \text{ for } i \leq m, && (40)\\ h_j(w) &\leq 0, \text{ for } j \leq n, && (41)\\ \alpha_j &\geq 0, \text{ for } j \leq n, && (42)\\ \alpha_j h_j(w) &= 0, \text{ for } j \leq n. && (43) \end{aligned}$$

However, in some cases, solving all these equations is intractable. An alternative is then to compute the dual problem,